This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Protein Engineering vol.1 no.4 pp.319–325, 1987

Improvement in the alkaline stability of subtilisin using an efficient random mutagenesis and screening procedure

Brian C.Cunningham and James A.Wells

Department of Biomolecular Chemistry, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080, USA

An efficient random mutagenesis procedure coupled to a replica plate screen facilitated the isolation of mutant subtilisins from *Bacillus amyloliquefaciens* that had altered autolytic stability under alkaline conditions. Out of about 4000 clones screened, approximately 70 produced subtilisins with reduced stability (negatives). Two clones produced a more stable subtilisin (positives) and were identified as having a single mutation, either Ile107Val or Lys213Arg (the wild-type amino acid is followed by the codon position and the mutant amino acid). One of the negative mutants, Met50Val, was at a site where other homologous subtilisins contained a Phe. When the Met50Phe mutation was introduced into the *B. amyloliquefaciens* gene, the mutant subtilisin was more alkaline stable. The double mutant (Ile107Val/Lys213Arg) was more stable than the isolated single mutant parents. The triple mutant (Met50Phe/Ile107Val/Lys213Arg) was even more stable than Ile107Val/Lys213Arg (up to two times the autolytic half-time of wild-type at pH 12). These studies demonstrate the feasibility for improving the alkaline stability of proteins by random mutagenesis and identifying potential sites where substitutions from homologous proteins can improve alkaline stability.

Key words: subtilisin/alkaline stability/mutagenesis

Introduction

The utility of industrial enzymes often depends on their stability in the presence of oxidants, organic solvents, extremes of pH and/or temperature. Subtilisin is a serine endopeptidase (for review see Markland and Smith, 1971) that is used industrially often in the presence of oxidants or under extremely alkaline conditions. Replacement of an oxidatively sensitive methionine residue by cassette mutagenesis (Wells *et al.*, 1985) of the cloned subtilisin gene from *Bacillus amyloliquefaciens* (Wells *et al.*, 1983) produced mutant subtilisins that are resistant to oxidative inactivation by hydrogen peroxide (Estell *et al.*, 1985). However, under prolonged incubation at high pH, subtilisin undergoes autolytic inactivation which may be related to the conformational stability of the molecule (Ottesen and Svendsen, 1970).

Protein conformational stability is determined by a complicated mixture of covalent and non-covalent binding forces. Our rudimentary understanding of these stabilizing interactions, and the typically small stabilization energy (5–15 kcal/mol) that distinguishes the folded and unfolded state(s) of proteins (for review see Pfeil, 1981; Creighton, 1983), currently pose a difficult challenge to the rational design of protein stability. While it has been possible to increase the thermal stability of T4 lysozyme to irreversible inactivation (Perry and Wetzel, 1984) and dihydrofolate reductase to reversible denaturation (Villafranca *et al.*, 1983) by introduction of disulfide bonds, this strategy has had mixed success in improving the thermal stability of subtilisin toward autolytic inactivation (Wells and Powers, 1986; Pantoliano *et al.*, 1987). Another rational approach to protein stability has been to introduce mutations that increase stability of secondary structural units such as α-helices (Hecht *et al.*, 1986; Mitchinson and Baldwin, 1986; Matthews *et al.*, 1987).

In contrast to the above site-specific approaches the most widely used strategy is to couple random mutagenesis with a suitable screen or selection to isolate point mutants that are more thermally stable. This strategy has been used to increase the thermal stability of the amino-terminal domain of phage lambda repressor (Hecht *et al.*, 1984), staphylococcal nuclease (Shortle and Lin, 1985; Shortle and Meeker, 1986; Shortle, 1986), T4 lysozyme (Alber and Wozniak, 1985) and kanamycin nucleotidyl transferase (Matsumura and Aiba, 1985; Liao *et al.*, 1986). In some cases these point mutants have cumulative effects on protein stability (Shortle, 1986; Shortle and Meeker, 1986; Matsumura *et al.*, 1986; Liao *et al.*, 1986; Hecht *et al.*, 1986).

Here we describe an improved random mutagenesis and screening procedure that facilitated the isolation of mutant subtilisins which are more autolytically stable (in a cumulative fashion) under alkaline conditions. Thus, the stability of this protein to extreme alkaline conditions is not optimal and can be increased by random mutagenesis.

Materials and methods

Reagents

Resolved Sp-diastereoisomers of deoxyguanosine 5′-O-(1-thiotriphosphate) (dGTPαS) and dATPαs and racemic mixtures of dTTPαS and dCTPαS were provided by Dr Phil Buzby (New England Nuclear). ATP and deoxynucleotide triphosphates were from PL Biochemicals. Dam methylase, S-adenosylmethionine, T4 DNA kinase, T4 DNA ligase and all restriction endonucleases were from New England Biolabs. DNA polymerase I large fragment (Klenow) and AMV polymerase were from Boehringer-Mannheim and Life Sciences Corp., respectively. Oligodeoxyribonucleotides were provided by the Organic Chemistry Group at Genentech.

Construction of random mutagenesis library

A convenient *E. coli–B. subtilis* shuttle vector, pBO180 (Figure 1) was constructed to contain the 2.3 kb *Eco*RI–*Pvu*II fragment comprising the *E. coli* 322 origin and ampicillin resistance gene (amp^r) from pBR327 (Covarrubias *et al.*, 1981), the 3.7 kb *Eco*RI–*Bam*HI fragment containing the UB110 origin, the chloramphenicol (cmp^r) and neomycin (neo^r) resistance genes from pBD64 (Gryczan *et al.*, 1980) and the 1.5 kb *Eco*RI–*Bam*HI subtilisin gene fragment (Wells *et al.*, 1983). In addition, a unique and silent *Kpn*I site at codon 166 was introduced into the subtilisin gene by site-directed mutagenesis (Zoller and Smith, 1982) to facilitate sub-cloning of mutant subtilisins. DNA fragments to be ligated were produced by restriction digestion of plasmid DNA (0.2 to 1 μg) and purified from 0.8% low gel temperature agarose (BioRad) in TAE buffer (Maniatis *et al.*, 1982) under sterile conditions. Gel slices containing relevant DNA fragments were melted by heating to 68°C for 5 min in 3.5 volumes of H_2O. DNA fragments were ligated in the agarose solution by

© IRL Press Limited, Oxford, England

incubation with T4 DNA ligase (100 units/ml final), ligase buffer (New England Biolabs) at 24°C for 1−16 h. Agarose gel ligation mixtures were added to two volumes of freshly competent *E. coli* LE392 cells (Enquist and Weisberg, 1977) for transformation according to Mandel and Higa (1970).

Deoxyuridine-containing template DNA from M13mp11 SUBT (Wells and Powers, 1986) was prepared according to Kunkel (1985), except a more stable *E. coli* strain (BO265) was used to produce DNA. BO265 was constructed by mating *E. coli* RZ1032 (tet^r; Kunkel, 1985) with K5303 (cmp^r, provided by Dr Harvey Miller, Genentech) which contains a chloramphenicol resistance gene on its F′ episome. A conjugate host (tet^r, cmp^r) was selected on LB plates containing 12.5 μg/ml chloramphenicol and 20 μg/ml tetracycline. Template DNA was purified by CsCl density gradients (Maniatis *et al.*, 1982).

A primer (*AvaI*$^-$) having the sequence 5′GAAAAAAGA-CCCTAGCGTCGCTTA3′ ending at codon −11 within the subtilisin gene, was used to alter a unique *AvaI* recognition sequence without changing the protein sequence. (The asterisk denotes the mismatch from the wild-type sequence and underlined is the altered *AvaI* site.) The 5′ phosphorylated *AvaI*$^-$ primer (~320 pmol) was annealed to the deoxyuridine containing M13mp11 SUBT template (~40 pmol) in 1.88 ml of 53 mM NaCl, 7.4 mM $MgCl_2$ and 7.4 mM Tris−HCl (pH 7.5) by heating to 90°C for 2 min and cooling to 24°C over 15 min (Figure 1). Primer extension was initiated at 24°C by addition of 100 units of Klenow fragment plus 50 μM of dATP, dTTP, dGTP and dCTP (final). Aliquots (50 μl) of the mixture were sampled every 15 s over 10 min and extension reactions were stopped by addition of 40 mM EDTA (final), pH 8.0. Pooled DNA samples were extracted with phenol/$CHCl_3$, precipitated twice with ethanol (Maniatis *et al.*, 1982), and redissolved in 0.4 ml 1 mM EDTA, 10 mM Tris−HCl, pH 8.

Separate misincorporation reactions for addition of each of the four α-thiodeoxynucleotides onto the 3′ ends of the randomly terminated template pool (~20 μg) were carried out by reaction with 0.25 mM of a given dNTPαs, 100 units AMV polymerase, 50 mM KCl, 10 mM $MgCl_2$, 0.4 mM dithiothreitol and 50 mM Tris (pH 8.3) in 0.2 ml total volume at 37°C for 90 min (Champoux, 1984). Extension from the site of misincorporation was performed by reaction with 50 μM all four dNTPs (pH 8.0), 50 units AMV polymerase at 37°C for 5 min. After ethanol precipitation, closed circular heteroduplexes were synthesized by reaction for 2 days at 14°C under the same conditions used for the timed extension reactions above except that the reactions contained 1000 units T4 DNA ligase, 0.5 mM ATP and 1 mM 2-mercaptoethanol.

Heteroduplex DNA in each reaction mixture was methylated by incubation with 80 μM S-adenosylmethionine and 150 units *dam* methylase for 1 h at 37°C. After heating at 68°C for 15 min, half of each of the four methylated heteroduplex reactions was transformed into 2.5 ml competent *E. coli* JM101 (Messing, 1979). The number of independent transformants from each of the four transformations ranged from 0.4 to 2.0 × 10^5. After growing up phage pools, RF DNA from each of the four transformations was isolated and purified by centrifugation through CsCl density gradients. Approximately 2 μg of RF DNA from each of the four pools was digested with *Eco*RI, *Bam*HI and *Ava*I. The 1.5 kb *Eco*RI−*Bam*HI fragment (i.e. *Ava*I resistant) was purified on low gel temperature agarose and ligated into the 5.5 kb *Eco*RI−*Bam*HI vector fragment of pBO180. The total number of independent transformants from each α-thiodeoxynucleotide misincorporation plasmid library ranged from 1.2 to 2.4 × 10^4. Plasmids from each of the four transformation pools were purified by centrifugation through CsCl density gradients.

Expression and screening of subtilisin point mutants

Pooled plasmid DNA was transformed (Anagnostopoulos and Spizizen, 1961) into BG2036, a strain of *B. subtilis* deficient in extracellular protease genes (Yang *et al.*, 1984). For each transformation, 5 μg of DNA produced approximately 2.5 × 10^5 independent BG2036 transformants. Fresh transformants were arrayed onto 96-well microtiter plates containing 150 μl per well LB media plus 12.5 μg/ml chloramphenicol. After 1 h at room temperature, a replica pattern was stamped (using a matched 96 prong replica stamp) onto a 132 mm BA 85 nitrocellulose filter (Schleicher and Scheull) which was layered on a 140 mm diameter agar plate containing LB media, 5 μg/ml chloramphenicol and 1.6% skim milk (Wells *et al.*, 1983). Cells were grown for about 16 h at 30°C until halos of proteolysis were roughly 5−7 mm in diameter. Filters were transferred directly to freshly prepared agar plates at 37°C containing 1.6% skim milk and 50 mM sodium phosphate pH 11.5, and incubated for 3−6 h at 37°C until wild-type subtilisin produced halos of about 5 mm. The plates were stained for 10 min at 24°C with Coomassie blue solution (0.25% Coomassie blue R-250 in 25% ethanol) and destained with 25% ethanol, 10% acetic acid for 20 min. Zones of proteolysis appeared as blue halos on a white background on the underside of the plate and were compared with the original LB-skim milk growth plate that was similarly stained and destained as a control. Clones were considered 'positive' that produced proportionately larger zones of proteolysis on the high pH plates relative to the original growth plate. 'Negative' clones gave smaller halos under alkaline conditions. Positive and negative clones were restreaked to colony purify and screened again in triplicate to confirm alkaline pH results.

Identification and analysis of mutant subtilisins

Plasmid DNA from 5 ml overnight cultures of BG2036 clones expressing mutant subtilisins was prepared according to Birnboim and Doly (1979) with minor modifications: cells were incubated with 2 mg/ml lysozyme for 5 min at 37°C to ensure cell lysis, and an additional extraction with phenol/$CHCl_3$ was employed to remove contaminants. The 1.5 kb *Eco*RI−*Bam*HI fragment containing the subtilisin gene was ligated into M13mp11 and template DNA was prepared for DNA sequencing (Messing and Vieira, 1982). Three DNA sequencing primers ending at codon −26, +95 and +155 were synthesized to match the subtilisin coding sequence. Single track DNA sequencing was used for preliminary identification of mutants. For example, a G sequence track was used to identify a mutant from the dGTPαs library. Four track DNA sequencing was performed over the site of mutagenesis to identify the mutant sequence (Sanger *et al.*, 1980).

Confirmed positive and negative bacilli clones were cultured in LB media containing 12.5 μg/ml chloramphenicol. Subtilisin was purified from culture supernatants as previously described (Estell *et al.*, 1985). Enzymes were greater than 98% pure as analyzed by SDS−polyacrylamide gel electrophoresis (Laemmli, 1970), and protein concentrations were calculated from the absorbance at 280 nm ($\epsilon_{280}^{0.1\%}$ = 1.17, Matsubara *et al.*, 1965). Enzyme activity was measured with 200 μg/ml succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide (Sigma) in 0.10 M Tris pH 8.6 or 0.10 M CAPS pH 10.8 at 25°C. Sp. act. (μmol product/min/mg) was calculated from the change in absorbance at 410 nm from production of *p*-nitroaniline (ϵ_{410} = 8480 M^{-1} cm^{-1}; Del Mar *et al.*, 1979). Alkaline autolytic stability studies were per-

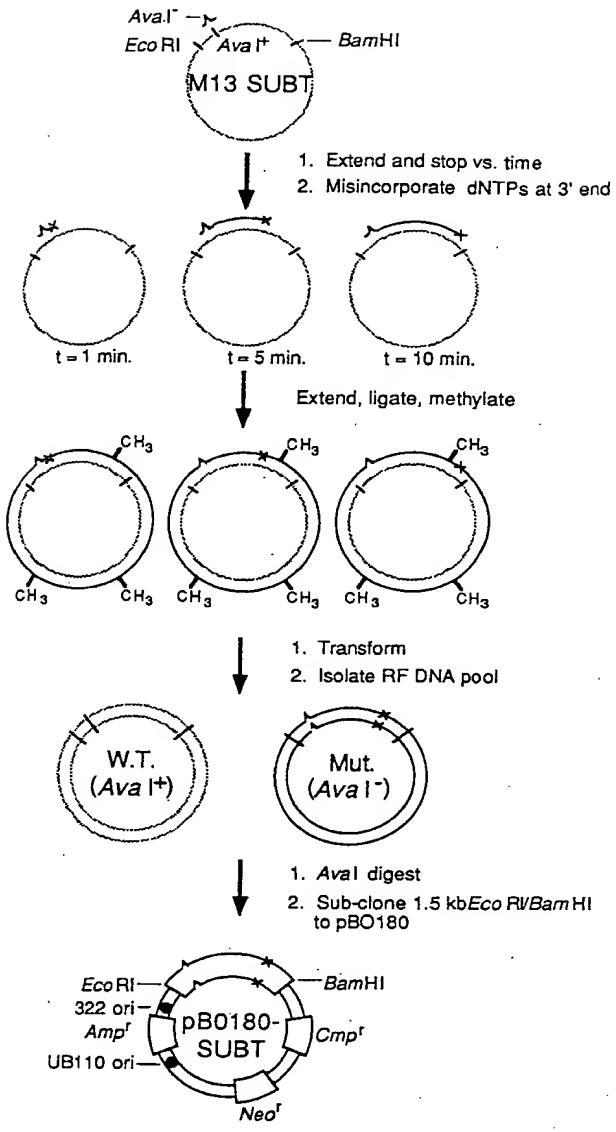

Fig. 1. Strategy for producing point mutations in the subtilisin coding sequence by misincorporation of α-thiodeoxynucleotide triphosphates. A restriction primer designed to eliminate a unique *AvaI* restriction site within the subtilisin gene was used to produce a set of randomly terminated primer template heteroduplexes (indicated by t = 1 min, 5 min and 10 min). α-Thiodeoxynucleotides were misincorporated (indicated by x) in four separate reactions. After heteroduplex synthesis, *in vitro* methylation and transformation, RF DNA pools from the four misincorporation reactions were isolated. *AvaI*-resistant inserts were subcloned into pBO180 for expression and screening in *B. subtilis* BG2036 (see Materials and methods for details).

formed on purified enzymes (200 μg/ml) in 0.10 M potassium phosphate (pH 12.0) at 25°C. At various times aliquots were assayed for residual enzyme activity (Wells and Powers, 1986).

Results

Optimization and analysis of mutagenesis frequency

A set of primer-template molecules that were randomly 3′-terminated over the mature subtilisin coding sequence (Figure 1) was produced by stopping polymerase reactions with EDTA after various times of extension from a fixed primer. The extent and distribution of duplex formation over the subtilisin gene fragment (1 kb) was assessed by multiple restriction digestion (not shown). For example, production of new *HinfI* fragments identified when polymerase extension had proceeded past Ile110, Leu233 and Asp259 in the subtilisin gene. The efficiency of each misincorporation reaction was estimated to be greater than 80% by the addition of each dNTPαs to the *AvaI* restriction primer, and by analysis from polyacrylamide gel electrophoresis (Hillebrand *et al.*, 1984). [Although it is possible to produce misincorporations with normal deoxynucleotides (Zakour *et al.*, 1984; Champoux, 1984), the α-thiophosphate linkage is resistant to exonucleolytic cleavage by *E. coli* DNA polymerase I (Shortle *et al.*, 1982). Therefore, we reasoned (but have not shown) that a misincorporated α-thiophosphate deoxynucleotide would be more stably retained *in vivo* compared to a normal phosphate deoxynucleotide].

Several manipulations were employed to maximize the yield of the mutant sequences in the heteroduplex. These included the use of a deoxyuridine containing template (Kunkel, 1985), *in vitro* methylation of the mutagenic strand (Pukkila *et al.*, 1983; Horton and Lord, 1986), and the use of *AvaI* restriction−selection against the wild-type template strand which contained a unique *AvaI* site. The separate contribution of each enrichment procedure to the final mutagenesis frequency was not determined, except that prior to *AvaI* restriction−selection ~30% of the segregated clones in each of the four pools still retained a wild-type *AvaI* site within the subtilisin gene. After *AvaI* restriction−selection >98% of the plasmids lacked the wild-type *AvaI* site. Subcloning of the *AvaI*-resistant *EcoRI*−*BamHI* subtilisin gene fragment by agarose gel purification, and *in situ* ligation with a similarly cut pBO180 vector fragment allowed large numbers of recombinants to be obtained (>100 000 per μg equivalent of input M13 DNA).

The frequency of mutagenesis for each of the four dNTPαs misincorporation reactions was estimated from the frequency that unique restriction sites were eliminated (Table I). Unique restriction sites chosen for this analysis within the subtilisin gene were *ClaI*, *PvuII* and *KpnI* located at codons 35, 104 and 166, respectively. As a negative control, the mutagenesis frequency was determined at the *PstI* site located in the β-lactamase gene which is outside the window of mutagenesis. Because the absolute mutagenesis frequency was close to the percentage of undigested plasmid DNA remaining after a single round of restriction digestion, two rounds of restriction−selection were necessary to reduce the background of surviving uncut wild-type plasmid below the mutant plasmid (Table I). The background of surviving plasmid from wild-type DNA probably represents the sum total of spontaneous mutations and residual undigested wild-type plasmid. Subtracting the frequency for unmutagenized DNA (background) from the frequency for mutant DNA, and normalizing for the window of mutagenesis sampled by a given restriction analysis (4−6 bp) provides a rough estimate of the mutagenesis efficiency over the entire subtilisin coding sequence (~1000 bp).

From this analysis, the average percentage of subtilisin gene inserts containing mutations that resulted from dGTPαs, dCTPαs or dTTPαs misincorporation was estimated to be 90, 70 and 20%, respectively. These high mutagenesis frequencies were generally quite variable depending upon the dNTPαs and misincorporation efficiencies at this site (Table I). For instance, misincorporation efficiency has been reported to be both dependent on the kind of mismatch, and the context of primer (Champoux, 1984; Skinner and Eperon, 1986). Furthermore, biased misincorporation efficiency of dGTPαs and dCTPαs over dTTPαs has been previously observed (Shortle and Lin, 1985). Unlike the dGTPαs, dCTPαs and dTTPαs libraries, the efficiency of mutagenesis for

Table I. Estimation of mutagenesis frequencies by restriction-site selection[a]

dNTPαs misincorporated[b]	Restriction site mutated	% resistant clones[c]			% resistant clones over background[d]	% mutants per 1000 bp[e]
		1st round	2nd round	Total		
None	*Pst*I	0.32	0.7	0.002	0	–
G	*Pst*I	0.33	1.0	0.003	0.001	0.2
T	*Pst*I	0.32	<0.5	<0.002	0	0
C	*Pst*I	0.43	3.0	0.013	0.001	3
None	*Cla*I	0.28	5	0.014	0	–
G	*Cla*I	2.26	85	1.92	1.91	380
T	*Cla*I	0.48	31	0.15	0.14	35
C	*Cla*I	0.55	15	0.08	0.066	17
None	*Pvu*II	0.08	29	0.023	0	–
G	*Pvu*II	0.41	90	0.37	0.35	88
T	*Pvu*II	0.10	67	0.067	0.044	9
C	*Pvu*II	0.76	53	0.40	0.38	95
None	*Kpn*I	0.41	3	0.012	0	–
G	*Kpn*I	0.98	35	0.34	0.33	83
T	*Kpn*I	0.36	15	0.054	0.042	8
C	*Kpn*I	1.47	26	0.38	0.37	93

[a]Mutagenesis frequency is estimated from the frequency for obtaining mutations at unique restriction sites within the mutagenized subtilisin gene (*Cla*I, *Pvu*II or *Kpn*I) compared with mutational frequency at a site outside the window of mutagenesis (*Pst*I).
[b]Plasmid DNA was from wild-type (none) or from a dNTPαs misincorporation library (G,T,C).
[c]Percentage of resistant clones was calculated from the fraction of clones remaining after digestion of the plasmid (greater than 3-fold excess digestion) with the indicated restriction enzyme and normalized to a non-digested control. Restriction-resistant plasmid DNA from the first round was subjected to a second round of restriction–selection. The total represents the product of the fractions of resistant clones obtained from both rounds of selection and is the percentage of restriction-site mutant clones in the original starting pool. Frequencies were derived from counting at least 20 and usually more than 100 colonies.
[d]Percent resistant clones was calculated by subtracting the percentage of restriction-resistant clones obtained for wild-type DNA (i.e. none) from that obtained for mutant DNA.
[e]The frequency of mutation over each restriction site was normalized to the number of possible bases (4–6 bp) within each restriction site that can be mutagenized by the misincorporation event, and extrapolated to the expected frequency over 1000 bp (the mature subtilisin gene).

the dATPαs misincorporation library could not be accurately assessed because 90% of the restriction-resistant plasmids analyzed simply lacked the subtilisin gene insert. The basis of this result is not clearly understood. However, correcting for this background, we estimate the mutagenesis frequency in the dATPαs misincorporation library at around 20%. In a separate experiment (not shown), the mutagenesis efficiencies for dGTPαs and dTTPαs misincorporation were estimated to be around 50 and 30%, respectively, based on the frequency of reversion of an inactivating mutation at codon 169.

The location of a mutation was readily determined by single track DNA sequencing over the entire gene that corresponded to the misincorporated α-thiodeoxynucleotide. The identity of the mutation was determined by four track DNA sequencing focused over the site of mutation. Of 14 mutants identified, the distribution was similar to that reported by Shortle and Lin (1985), except that we did not observe nucleotide insertion or deletion mutations. There was a bias for A → G mutations in the G misincorporation library, and some unexpected point mutations appeared in the dTTPαs and dCTPαs libraries.

Screening and identification of alkaline stability mutants of subtilisin

Two problems were posed by screening colonies under high alkaline conditions (>pH 11). Because *B. subtilis* will not grow at high pH, it was necessary to grow colonies on filters at neutral pH to produce subtilisin, and subsequently transfer filters to skim milk plates at pH 11.5 to assay subtilisin activity (Wells *et al.*, 1983). However, at pH 11.5 the casein micelles no longer formed a turbid background which prevented a clear observation of proteolysis halos. This problem was overcome by briefly staining the plate with Coomassie brilliant blue (R-250) to amplify proteolysis zones and acidifying the plates to develop casein micelle turbidity. By comparing the halo size produced on the reference growth plate (pH 7.0) with the high pH plate (pH 11.5), it was possible to identify mutant subtilisins that had increased (positives) or decreased (negatives) stabilities under alkaline conditions (not shown).

Approximately 1000 colonies were screened from each of the four misincorporation libraries. The percentages of colonies showing a differential loss of activity at pH 11.5 versus pH 7 were 1.4, 1.8, 1.4 and 0.6% of the colonies that expressed subtilisin activity at neutral pH from the dGTPαs, dATPαs, dTTPαs and dCTPαs libraries, respectively. Several of these negative clones were sequenced and all were found to contain a single base change as expected from the misincorporation library from which they came. Negative mutants included Asp36Ala, Lys170Glu and Met50Val (the wild-type amino acid is followed by the codon position and the mutant amino acid). Two positive mutants were identified as Ile107Val and Lys213Arg. The ratio of negatives to positives was roughly 50:1.

Stability and activity of subtilisin mutants at alkaline pH

Subtilisin mutants were purified and their autolytic stabilities were measured by the time course of inactivation at pH 12.0 (Figure 2, Table II). At the termination of each autolysis study, SDS–PAGE analysis confirmed (by disappearance of the 27.5 kd subtilisin band) that the subtilisin variant had autolyzed to an extent consistent with the remaining enzyme activity. Positive mutants identified from the screen (Ile107Val and Lys213Arg) were more resistant to alkaline induced autolytic inactivation compared with wild-type; negative mutants (Lys170Glu and Met50Val) were less

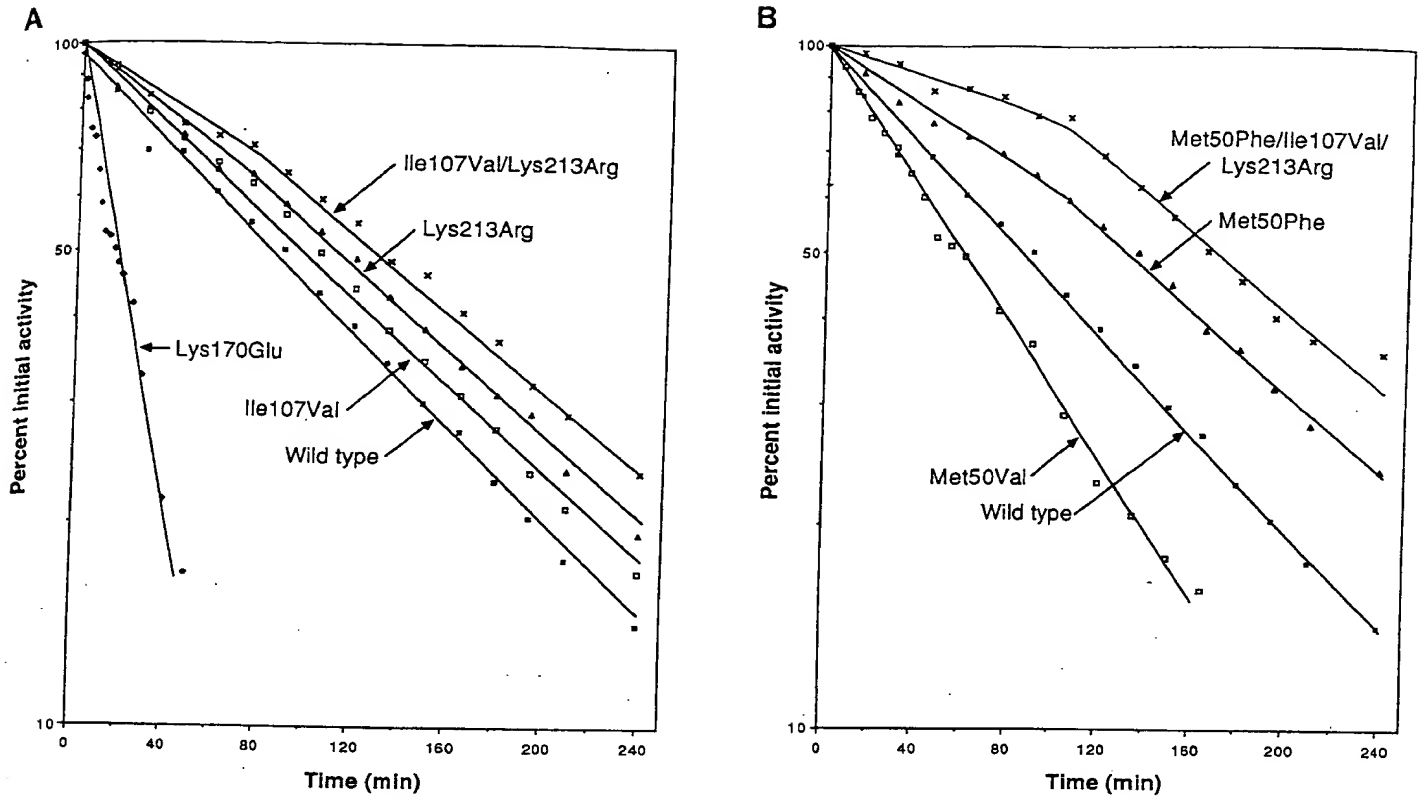

Fig. 2. Autolytic stability of purified wild-type and mutant subtilisins (200 μg/ml) in 100 mM potassium phosphate pH 12 at 25°C. After incubation for indicated times, residual enzyme activity was measured as described in Materials and methods. **Panel A**: wild-type subtilisin (■—■); Ile107Val (□—□); Lys213Arg (△—△); Lys170Glu (◇—◇); Ile107Val/Lys213Arg (x—x). **Panel B**: wild-type subtilisin (■—■); Met50Val (□—□); Met50Phe (△—△); Met50Phe/Ile107Val/Lys213Arg (x—x).

resistant. A Met50Phe mutant produced by site-directed mutagenesis (D.Powers and J.A.W., unpublished) was more stable than wild-type enzyme to alkaline autolytic inactivation (Figure 2B).

The stabilizing effects of Ile107Val, Lys213Arg and Met50Phe were cumulative as determined from rates of alkaline autolysis (Figure 2). The double mutant, Ile107Val/Lys213Arg, was more stable than either single mutant. The triple mutant, Met50Phe/Ile107Val/Lys213Arg, was more stable than the double mutant or Met50Phe. The inactivation curves showed a biphasic character that became more pronounced the more stable the mutant analyzed. This may have resulted from some destabilizing chemical modification(s) (e.g. deamidation) during the autolysis study and/or reduced stabilization caused by complete digestion of larger autolysis peptides. These alkaline autolysis studies have been repeated on separately purified enzyme batches with essentially the same results.

Rates of autolysis may depend both on the specific activity as well as the conformational stability of the subtilisin variant (Wells and Powers, 1986). It was therefore possible that the decrease in autolytic inactivation rates resulted from decreases in specific activity of the more apparently stable mutants under alkaline conditions. However, the more stable mutants, if anything, had relatively higher specific activities than wild-type under alkaline conditions and the less stable mutants have relatively lower specific activities (Table II). These subtle effects on specific activity for Ile107Val/Lys213Arg and Met50Phe/Ile107Val/Lys213Arg are cumulative at both pH 8.6 and 10.8. The changes in specific activity may reflect slight differences in substrate specificity, although only positions 170 and 107 are near the substrate binding site (Robertus *et al.*, 1972).

Discussion

Variable polymerase extension from a fixed primer permitted synthesis *in vitro* of a uniform set of 3′-terminated ends onto which misincorporation events could be focused on the subtilisin gene. This approach is in contrast to whole plasmid mutagenesis, where a single-stranded nick is produced randomly by treatment with DNase I in the presence of ethidium bromide, and the 3′ ter-

Table II. Relationship between relative specific activity at pH 8.6 or 10.8 and alkaline autolytic stability

Enzyme	Relative sp. act.[a]		Alkaline autolysis half-time (min)[b]
	pH 8.6	pH 10.8	
Wild-type	100 ± 1	100 ± 3	86
Lys170Glu	46 ± 1	28 ± 2	13
Ile107Val	126 ± 3	99 ± 5	102
Lys213Arg	97 ± 1	102 ± 1	115
Ile107Val/Lys213Arg	116 ± 2	106 ± 3	130
Met50Val	66 ± 4	61 ± 1	58
Met50Phe	123 ± 3	157 ± 7	131
Met50Phe/Ile107Val/Lys213Arg	126 ± 2	152 ± 3	168

[a]Relative sp. act. was the average from triplicate activity determinations normalized to the wild-type subtilisin value at the same pH. The average sp. act. of wild-type enzyme at pH 8.6 and 10.8 was 70 μmol/min/mg and 37 μmol/min/mg, respectively.
[b]Time to reach 50% activity was taken from Figure 2.

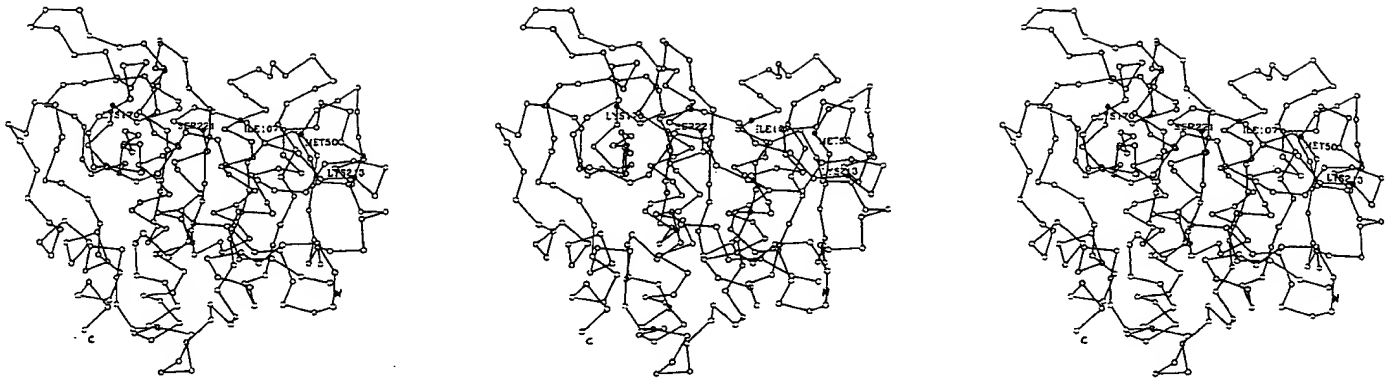

Fig. 3. Stereoview of the α-carbon diagram of subtilisin showing the positions of more alkaline stable mutants (Met50Phe, Ile107Val and Lys213Arg), less alkaline stable mutants (Met50Val and Lys170Glu) and the catalytic Ser221.

minus is exposed for misincorporation by limited digestion with exonuclease III (Shortle *et al.*, 1982). The former method is necessarily more regionally efficient because the misincorporation events are concentrated on the segment of interest instead of being diluted over the entire plasmid. Furthermore, the wild-type background arising from plasmid that is not efficiently nicked with DNase I is avoided by starting with a single-stranded template such as M13. Other groups have utilized φX174 (Zakour *et al.*, 1984) or M13 (Champoux, 1984; Kunkel, 1985; Skinner and Eperon, 1986; Singh *et al.*, 1986) templates to produce mutations by misincorporation immediately behind or by fixed extension from a synthetic oligonucleotide primer instead of behind a randomly extended primer as employed here. The use of enrichment procedures such as deoxyuridine containing DNA (Kunkel, 1985), *in vitro* methylation (Pukkila *et al.*, 1983; Horton and Lord, 1986) and restriction-primer selection produced high frequencies of mutagenesis over the subtilisin gene fragment, (~ 1 kb).

Uncoupling the growth of cells from the high pH screen circumvented the problem that the *B. subtilis* cells will not grow at pH 11.5. Use of a general protease substrate casein, instead of a specific synthetic substrate, avoided selecting effects of pH on substrate specificity. In principle, the alkaline screen should be capable of identifying mutants having both greater alkaline activity as well as alkaline stability. In fact, the mutants identified as positive were all more stable and somewhat more alkaline active than wild-type. This may explain why the zones of proteolysis under alkaline conditions (not shown) for the positive mutants appeared more pronounced relative to wild-type than expected from the autolytic stability under alkaline conditions measured *in vitro*.

Although the number of mutants characterized is too small to generalize, a number of points are noteworthy. All mutants described are located on, or are accessible to, the surface of the molecule (Figure 3). Alber *et al.* (1987) have shown that random mutations in buried positions of T4 lysozyme are more likely to destabilize the protein compared to mutations at surface-accessible sites. We probably missed mutant enzymes which were extremely unstable, because we chose to screen only colonies expressing active subtilisins at neutral pH. Position 50 is located at the end of a β sheet structure. The Met side chain makes van der Waals contact with hydrophobic side chains of Trp113, Val93 and Val95 (R.Bott and M.Ultsch, unpublished results). Modelling of the Met50Val substitution shows a β-branched constituent would make unfavourable steric contacts with main chain carbonyl oxygens at positions 94 and/or 106. Model building of a Met50Phe substitution shows the Phe side chain can make a potentially more favourable hydrophobic interaction with Trp113. Such an interaction is observed between Phe50 and Trp113 in the X-ray crystal structure of the *B. licheniformis* subtilisin (McPhalen *et al.*, 1985). Improved packing interactions between aromatic side chains can confer added stabilization to proteins (Burley and Petsko, 1985).

Position 170 is located in a loop that forms part of the P1 binding site of subtilisin (Robertus *et al.*, 1972). The Lys side chain is close to Ser165, Tyr167, and is 4 Å from Glu195 (NZ-170 to OE2-195). The Lys170Glu mutant probably causes unfavorable electrostatic effects at high pH where the neighbouring side chains from Glu195 and Tyr167 would also be negatively charged. Ile107 is an α-helix that comprises part of the P4 binding site (Robertus *et al.*, 1972). The δ-methyl group is not in contact with other atoms from the enzyme and it is unclear why a Val substitution should marginally improve alkaline autolytic stability. Lys213 is near the end of a β-ribbon (Figure 3). The side chain is accessible to solvent and the X-ray structure indicates that it does not make any intramolecular electrostatic interactions. Models of Lys213Arg do not suggest the possibility of new electrostatic contacts. It has been shown that enzymes can be stabilized by guanidination of lysine side chains (Cupo *et al.*, 1980). In addition, we expect that at pH 12 the Lys side chain would be neutral whereas Arg would be at least partially positively charged (Tanford, 1962).

Autolysis of subtilisin is promoted by agents which disrupt the conformational integrity of the molecule such as high temperature, denaturants, high pH, and chelants which remove stabilizing calcium ions (Ottesen and Svendsen, 1970; Voordouw *et al.*, 1976; Wells and Powers, 1986). It is possible that mutations which stabilize the molecule to alkaline autolytic inactivation may not effect thermal autolytic inactivation. In fact, preliminary experiments indicate that the alkaline stable triple mutant, Met50Phe/Ile107Val/Lys213Arg, is not significantly more autolytically stable than wild-type subtilisin at 55°C in 50 mM MOPS (pH 7), 50 mM NaCl and 2 mM $CaCl_2$ (B.C.C. and J.A.W., unpublished).

The single positive mutants identified are not more dramatically stable than the wild-type. However, the fact that the stabilizing effects from these positions are cumulative permits large improvements in alkaline autolytic stability to be achieved. In addition, the stabilizing mutation (Met50Phe) is at a site where a destabilizing mutation has been identified from the random mutagenesis (Met50Val). The Met50Phe substitution is found in other bacillus subtilisins sequenced (Markland and Smith, 1971; Nedkov *et al.*,

1983; Stahl and Ferrari, 1984). Thus, a strategy to improve alkaline autolytic stability further can involve producing substitutions (especially variant substitutions from natural extremophiles) at sites found by random mutagenesis to be alkaline (and perhaps thermally) sensitive such as was done at position 50. While this manuscript was in press, a more thermally stable variant of subtilisin from *B. amyloliquefaciens* (Asn218Ser) was isolated by random mutagenesis (Bryan *et al.*, 1987). A serine at position 218 is found in natural variant subtilisins from thermophilic sources. These studies further demonstrate the usefulness of the random mutagenesis and screening approach for improving the range and utility of enzymes for industrial purposes.

Acknowledgements

We are grateful to Phillip Kiboneka for preparation of media and plates, the organic chemistry group at Genentech for oligonucleotide synthesis, Wayne Anstine for preparation of the manuscript, Ron Rubin for help with graphics, Paul Carter and Robert Caldwell for critical reading of the manuscript, and colleagues at Genentech and Genencor for their support and encouragement.

References

Alber,T. and Wozniak,J.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 747–750.
Alber,T., Dao-Pin,S., Nye,J.A., Muchmore,D.C. and Matthews,B.W. (1987) *Biochemistry*, **26**, 3754–3758.
Anagnostopoulos,C. and Spizizen,J. (1961) *J. Bacteriol.*, **81**, 741–746.
Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513–1528.
Bryan,P.N., Rollence,M.L., Pantoliano,N.W., Wood,J., Finzel,B.C., Gilliland,G.L., Howard,A.J. and Poulos,T.L. (1987) *Proteins*, **1**, 326–334.
Burley,S.K. and Petsko,G.A. (1985) *Science*, **229**, 23–28.
Carter,P., Bedouelle,H. and Winter,G. (1985) *Nucleic Acids Res.*, **13**, 4431–4443.
Champoux,J.J. (1984) *J. Mol. Appl. Genet.*, **2**, 454–464.
Covarrubias,L., Cervantes,L., Covarrubias,A., Soberon,X., Vichido,J., Blanco,A., Portnoy,Y.K. and Bolivar,F. (1981) *Gene*, **13**, 25–35.
Creighton,T.E. (1983) In *Proteins: Structure and Molecular Properties*. Freeman, NY, Ch. 7.
Cupo,P., El-Deiry,W., Whitney,P.L. and Awad,W.M. (1980) *J. Biol. Chem.*, **255**, 10828–10833.
Del Mar,E.G., Largman,C., Brodrick,J.W. and Goekas,M.C. (1979) *Anal. Biochem.*, **99**, 316–320.
Dente,L., Cesareni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645–1655.
Enquist,L.W. and Weisberg,R.A. (1977) *J. Mol. Biol.*, **111**, 97–120.
Estell,D.A., Graycar,T.P. and Wells,J.A. (1985) *J. Biol. Chem.*, **260**, 6518–6521.
Gryczan,T., Shivakumar,A.G. and Dubnau,D. (1980) *J. Bacteriol.*, **141**, 246–253.
Hecht,M.H., Sturtevant,J.M. and Sauer,R.T. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5685–5689.
Hecht,M.H., Sturtevant,J.M. and Sauer,R.T. (1986) *Proteins*, **1**, 43–46.
Hillebrand,G.G., McCluskey,A.H., Abbott,K.A., Revich,G:G. and Beattie,K.L. (1984) *Nucleic Acids Res.*, **12**, 3155–3171.
Horton,R.E. and Lord,S.T. (1986) *Nucleic Acids Res.*, **14**, 5112.
Kramer,W., Schughart,K. and Fritz,H.J. (1982) *Nucleic Acids Res.*, **10**, 6475–6485.
Kunkel,T.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 488–492.
Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
Liao,H., McKenzie,T. and Hageman,R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 576–580.
Mandel,M. and Higa,A. (1970) *J. Mol. Biol.*, **53**, 159–162.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (eds) (1982). In *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
Markland,F.S. and Smith,E.L. (1971) In Boyer,P.D. (ed.), *The Enzymes*. Academic Press, NY, Vol. 3, pp. 561–608.
Matsubara,H., Kasper,C.B., Brown,D.M. and Smith,E.L. (1965) *J. Biol. Chem.*, **240**, 1125–1130.
Matsumura,M. and Aiba,S. (1985) *J. Biol. Chem.*, **260**, 15298–15303.
Matsumura,M., Yasumura,S. and Aiba,S. (1986) *Nature*, **323**, 356–358.
Matthews,B.W., Nicholson,H. and Becktel,W.J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, in press.
McPhalen,C.A., Schnebli,H.P. and James,M.N.G. (1986) *FEBS Lett.*, **188**, 55–88.
Messing,J. (1979) *Recombinant DNA Tech. Bull.*, **2**, 43–48.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269–276.
Mitchinson,C. and Baldwin,R.L. (1986) *Proteins*, **1**, 23–33.
Nedkov,P., Oberthür,W. and Braunitzer,G. (1983) *Hoppe-Seyler's Z. Physiol. Chem.*, **364**, 1537–1540.
Ottesen,M. and Svendsen,I. (1970) *Methods Enzymol.*, **19**, 199–215.
Pantoliano,M.W., Ladner,R.C., Bryan,P.N., Rollence,M.L., Wood,J.F. and Poulos,T.L. (1987) *Biochemistry*, **26**, 2077–2082.
Perry,L.J. and Wetzel,R. (1984) *Science*, **226**, 555–557.
Pfeil,W. (1981) *Mol. Cell. Biochem.*, **40**, 3–28.
Pukkila,P.J., Peterson,J., Herman,G., Modrich,P. and Meselson,M. (1983) *Genetics*, **104**, 571–582.
Robertus,J.D., Alden,R.A., Birktoft,J.J., Kraut,J., Powers,J.C. and Wilcox,P.E. (1972) *Biochemistry*, **11**, 2439–2449.
Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.M. and Rue,B.A. (1980) *J. Mol. Biol.*, **143**, 161–178.
Shortle,D. (1986) *J. Cell. Biochem.*, **30**, 281–289.
Shortle,D. and Lin,B. (1985) *Genetics*, **110**, 539–555.
Shortle,D. and Meeker,A.K. (1986) *Proteins: Structure, Function, and Genetics*, **1**, 81–89.
Shortle,D., Grisafi,P., Benkovic,S.J. and Botstein,D. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 1588–1592.
Singh,M., Heaphy,S. and Gait,M.J. (1986) *Protein Engineering*, **1**, 75–76.
Skinner,J.A. and Eperon,I.C. (1986) *Nucleic Acids Res.*, **14**, 6945–6964.
Stahl,M.L. and Ferrari,E. (1984) *J. Bacteriol.*, **158**, 411–418.
Tanford,C. (1962) *Adv. Protein Chem.*, **17**, 69–165.
Villafranca,J.A., Howell,E.E., Voet,D.H., Strober,M.S., Ogden,R.C., Abelson, J.N. and Kraut,J. (1983) *Science*, **222**, 782–788.
Voordouw,R., Milo,C. and Roche,R.S. (1976) *Biochemistry*, **15**, 3716–3724.
Wells,J.A. and Powers,D.B. (1986) *J. Biol. Chem.*, **261**, 6564–6570.
Wells,J.A., Cunningham,B.C., Graycar,T.P. and Estell,D.A. (1986) *Phil. Trans. R. Soc. Lond: A*, **317**, 415–423.
Wells,J.A., Ferrari,E., Henner,D.J., Estell,D.A. and Chen,E.Y. (1983) *Nucleic Acids Res.*, **11**, 7911–7925.
Wells,J.A., Vasser,M. and Powers,D.B. (1985) *Gene*, **34**, 315–323.
Yang,M.Y., Ferrari,E. and Henner,D.J. (1984) *J. Bacteriol.*, **160**, 15–21.
Zakour,R.A. and Loeb,L.A. (1982) *Nature*, **295**, 708–710.
Zakour,R.A., James,E.A. and Loeb,L.A. (1984) *Nucleic Acids Res.*, **12**, 6615–6628.
Zoller,M.J. and Smith,M. (1982) *Nucleic Acids Res.*, **10**, 6487–6500.

Received on June 3, 1987

Protein Engineering vol.3 no.3 pp.181–191, 1990

Random mutagenesis used to probe the structure and function of *Bacillus stearothermophilus* alpha-amylase

Liisa Holm, Anu K.Koivula, Päivi M.Lehtovaara, Ari Hemminki and Jonathan K.C.Knowles

Biotechnical Laboratory, VTT, Tietotie 2, SF-02150 Espoo, Finland

Mutations that cover the sequence of *Bacillus stearothermophilus* α-amylase were produced by an efficient *in vitro* enzymatic random mutagenesis method and the mutant α-amylases were expressed in *Escherichia coli*, which also secreted the product. Ninety-eight mutants were identified by sequencing and their enzyme activities were classified into three classes: wild-type, reduced or null. A molecular model of the enzyme was constructed using the coordinates of Taka-amylase A and a consensus alignment of mammalian, plant, and bacterial α-amylases. The location of mutant amino acids on the model indicate that mutations which destroy or decrease the catalytic activity are particularly clustered: (i) around the active site and along the substrate-binding groove and (ii) in the interface between the central α/β barrel and the C-terminal domain. Exposed loops are typically tolerant towards mutations.

Key words: homology modelling/sequence alignment/starch

Introduction

α-Amylases (α-1,4-glucan-4-glucanohydrolase, EC 3.2.1.1) are widely distributed starch-degrading enzymes found in plants, animals and bacteria. Because of the importance of starch as a raw material in a number of industries, particularly the food industries, α-amylase is an enzyme of considerable commercial significance. Despite this, there is relatively little information concerning either structure or function of this interesting family of enzymes. It is a common observation that even the availability of a good three-dimensional structure does not necessarily make it possible to predict the effects of specific amino acid changes, even in the case where information about the enzyme mechanism is available (Knowles, 1988). For this reason we have developed a novel random mutagenesis method (Lehtovaara *et al.*, 1988). The method permits the enzymatic generation of a library of single (and multiple) base mutations throughout a defined region of DNA which can be several kilobases long. This method has now been applied to the cloned gene of *Bacillus stearothermophilus* α-amylase. A large number of mutations have been identified by sequencing in the course of refining the methodology.

The three-dimensional structures of two α-amylases, one from *Aspergillus oryzae*, the so-called Taka-amylase A, and the other from pig pancreas, have been determined by X-ray crystallography (Matsuura *et al.*, 1984; Buisson *et al.*, 1988). Both α-amylases have three domains: A, B and C. Overall, their structures appear very similar, but some differences are observed in the structure of domain B and in the orientation of domain C relative to domain A (Buisson *et al.*, 1988). The N-terminal domain A corresponds to the well-known $(\alpha/\beta)_8$ barrel motif, which was first seen in triose phosphate isomerase (Phillips *et al.*, 1978). In the middle of domain A, the chain loops out to make domain B, which forms a lid above the barrel. The active site is located at the top of the barrel on the C-terminal side of the parallel β-sheet. Substrate analogues bind in this site (Matsuura *et al.*, 1984; Buisson *et al.*, 1988), and there are three carboxylic acids around this site which are conserved in all known α-amylase sequences. In *B.stearothermophilus* α-amylase they correspond to residues D234, E264 and D331. Matsuura *et al.* (1984) propose that residues E231 and D297 of Taka-amylase A, corresponding to D331 and E264 of *B.stearothermophilus*, are the catalytic residues analogously to the mechanism in lysozyme. Buisson *et al.* (1988) propose that residues D196 and D300 of porcine pancreatic α-amylase, corresponding to D234 and D331 of *B.stearothermophilus*, are catalytic. A binding site for calcium, which is essential for activity (Steer and Levitzky, 1973), is located between domains A and B. The C-terminal region is folded into an antiparallel β-barrel and forms domain C. Although it has proved possible to crystallize the enzyme from *B.stearothermophilus* (Ogasahara *et al.*, 1970), X-ray structures for the bacterial α-amylases are not yet available.

It has been shown that tertiary structures are better conserved in evolution than amino acid sequences (Chothia and Lesk, 1986). Computer-aided molecular modelling has been used in a number of cases to construct a model of a protein of unknown structure which is related to a protein with a known structure [for a review see Blundell *et al.* (1987)]. α-Amylase genes from a variety of sources have been sequenced. Limited inter-species homology has been reported to be located in short stretches around the active site (Svensson, 1988). Using a sensitive method, we have found sufficient homology over almost all of the amino acid sequence which justifies homology modelling of the *B.stearothermophilus* sequence onto the Taka-amylase A structure. It is known that the three-dimensional structures of porcine pancreatic α-amylase and Taka-amylase A are similar. We show that the sequence homology between *B.stearothermophilus* and Taka-amylase A is more significant than between the former pair. The Taka-amylase A structure was chosen because it is currently the only publicly available α-amylase structure.

In this paper, we present a three-dimensional structural model of *B.stearothermophilus* α-amylase based on the structure of Taka-amylase A and show that the properties of the random mutants can be consistently interpreted with the aid of this model.

Materials and methods

Generation of point mutations

The α-amylase gene from *B.stearothermophilus* ATCC 12980 was cloned as a 1.9 kb long fragment to the phage vectors M13mp18 (coding strand) and M13mp19 (non-coding strand). The insert contains the complete coding region of the α-amylase gene (1650 bp). Random enzymatic mutagenesis of the cloned DNA was performed according to the method of Lehtovaara *et al.* (1988). The method can be summarized as follows. Starting from oligonucleotide primers, a population of molecules is synthesized under conditions where one nucleotide is limiting. This gives a population of molecules each of which ends just

© Oxford University Press

before the omitted base. After removal of all nucleotides, misincorporation with reverse transcriptase is carried out under optimized conditions to give, in principle, a 1:1:1 ratio of the incorrect bases. Following synthesis of the second strand, the mutated dsDNA is transfected into *Escherichia coli*. The mutant library was generated in the course of developing the random mutagenesis method. FIve oligonucleotides (20-mers) at ~200 nucleotide intervals were used to mutagenize the A nucleotides in the coding region of α-amylase in the M13mp19 construction. With one of these primers the template nucleotides G and C were also mutagenized. Only one primer was used to mutagenize T and G nucleotides in the M13mp18 construction (covering amino acids H208–L248).

Characterization of mutants

A suitable *dut*$^+$*ung*$^+$ strain of *E.coli*, such as JM109 or TG2, was transfected with the mutagenized DNA and the cells were plated on glucose minimal medium/H-top amylopectin azure plates. Expression of functional recombinant α-amylase in *E.coli* was detected as halos around the plaques. A three-grade scale was used to assay the α-amylase activity as follows: ++ for wild-type-size halos, + for smaller but visible halos and − for non-detectable halos. Mutations were identified by DNA sequencing.

Atomic coordinates for the starting model

The atomic coordinates for the starting model were from the structure of Taka-amylase A at 3.0 Å resolution [Matsuura *et al.*, 1984; available from the Protein Data Bank (Bernstein *et al.*, 1977) as entry 2TAA].

Sequence alignment

Because of a high level of noise, automatic Needleman–Wunsch (1970) procedures are not applicable to the alignment of sequences that are only ~20% identical, which is the case for most sequence pairs in our sample. The sensitive alignment program of Argos (1987) was used for initial pairwise alignments. Discrepancies between different pairwise alignments were resolved by visual pattern matching to produce a consensus alignment of the sequences. The positions of α helices and β strands were taken for pig pancreatic α-amylase from Buisson *et al.* (1988), and for Taka-amylase A they were calculated from the coordinates using the DSSP program (Kabsch and Sander, 1983). Helices and strands of pig pancreatic α-amylase and Taka-amylase A were matched. Gaps were introduced to maximize similarity across all sequences according to the following conservative substitution groups: {P, G}, {S, T}, {E, D, N, Q}, {H, Y, F, W}, {A, I, L, V, M, C}, {K, R}. At positions corresponding to a known helix or strand, however, gaps were not allowed in any sequence. The significance of pairwise alignments was assessed by the correlation of five physical characteristics of amino acids (hydrophobicity, bulkiness, turn preference, strand preference, refractivity index) as given by Argos (1987). This measure proved quite sensitive to small changes such as the movement of just a single residue in a gap. Argos (1987) recommends that alignments where the average correlation is at >2.0 SD above random be considered seriously. Adjustments were made to the alignment until the value was about this threshold for the majority of pairwise alignments.

Model building

It was assumed that the backbone structure in aligned positions is the same between Taka-amylase A and *B.stearothermophilus* α-amylase. Loops involving gaps were patched using fragments selected from the database of known structures (Jones and Thirup, 1986). One to three C(α) atoms at both ends of the loop were matched to the template (derived from Taka-amylase A). Different anchors were tried out until a fragment was found where the fit to C(α)s was good (deviation of superposed fragment to anchor atoms <1 Å). Ideally, all main chain atoms of the fragment and of the starting structure would overlap, but in practice the overlap could be restricted to just one residue at both ends. Main chain coordinates were replaced by those of the loop starting from the second residue in the N-side anchor region and ending with the second-last residue in the C-side anchor region. Graphics was used to make sure that the selected loops packed tightly to the rest of the structure. Structures that would clash with the rest of the molecule were rejected already when they were collected from the database. Side chain coordinates for aligned identical residues were copied from the Taka-amylase A structure, otherwise they were inserted in standard conformation. The conformation of F346 and W492 was changed to another staggered conformation because they initially clashed with the main chain.

Strain introduced to the model through sidechain substitutions and joining of loops was relaxed by the program CHARMm (Brooks *et al.*, 1983) using an adapted basis Newton–Raphson energy minimization with a distance-dependent dielectric constant and an 8 Å cut off for non-bonded interactions. The calcium ion was not included in the minimization.

A chain of six glucose units joined by α-1,4-linkages was built in the interactive molecular modelling program Quanta (Polygen Inc.). The dihedral angles in the aldose linkage were +90 and −31° for O5-C1-O4-C4 and C1-O4-C4-C3 respectively. The interaction energy of a water probe to protein was calculated in Quanta. Energy countours at an arbitrarily chosen level of −2.0 kcal/mol formed a tube which passed by the catalytic residues, and the glucose chain was fitted manually to it. Rotation of one glycosidic bond was necessary to fit the chain in contact with the catalytic residues (D234, E264, D331). The substrate was relaxed by energy minimization keeping the protein part frozen.

Results

Alignment of α-amylase sequences

The alignment of distantly related α-amylase sequences is far from obvious because only six blocks of residues are clearly conserved in them (Svensson, 1988). However, when the correlation of physical properties of amino acids is used as a measure, an acceptable alignment in terms of standard deviations above random sequence comparison scores can be made over the whole length of the sequences (Table I). The alignment in Figure 1 shows that the best conserved regions are in domains A and C, and these are segments around the catalytic residues and secondary structure elements where particularly hydrophobic patches are conserved. Although domain C lacks invariant residues, the patterns of alternating hydrophobic and hydrophilic residues are preserved in the β strands.

In the aligned α-amylase sequences (Figure 1) there are 31 positions where the nature of the amino acid is conserved. Of these, 13 positions require hydrophobic residues, three are invariantly glycine, three require an aromatic residue and seven require an invariant charged residue.

In domain B there is very little amino acid homology. Further, the arrangement of β strands in the two known structures, porcine pancreatic α-amylase and Taka-amylase A, is distinctly different (Buisson *et al.*, 1988). In addition, in the amyloliquefying

Table I. Quality of the alignment

Sequence	1	2	3	4	5	6	7	8	9	10
1 *S.hygroscopicus*		8.4	3.9	1.4	0.7	3.0	0.3	1.3	1.4	−0.1
2 Pig pancreas	38		5.0	1.2	1.4	2.8	1.6	2.1	2.3	1.7
3 *B.subtilis*	24	24		2.3	2.1	2.7	0.5	1.5	2.1	1.8
4 *B.circulans*	21	18	21		4.4	5.5	1.3	1.3	1.5	0.8
5 *S.fibuligera*	19	10	21	28		14.5	3.0	3.3	3.5	1.1
6 Taka-amylase	22	23	20	27	52		3.7	3.2	3.4	1.0
7 *B.stearothermophilus*	20	16	19	22	23	24		17.7	17.1	2.2
8 *B.licheniformis*	18	16	21	20	23	21	66		20.3	3.2
9 *B.amyloliquefaciens*	19	17	21	21	23	23	66	81		3.7
10 Barley	17	17	18	18	19	20	24	25	26	

Upper part of the table: number of SD above random for the alignment in Figure 1 over five parameters as given by the sensitive homology program of Argos (1987). If 2.5 standard deviations is used as a criterion, all but barley are satisfactorily aligned with Taka-amylase A. The structure of Taka-amylase A (*Aspergillus oryzae*) was used as a template for the structural model of *B.stearothermophilus*. Lower part of the table: percentage of identical amino acids of aligned positions, i.e. positions not occupied by a gap in either sequence of a pairwise comparison. Aligned positions account for at least 70% of the length of any sequence in any pairwise comparison.

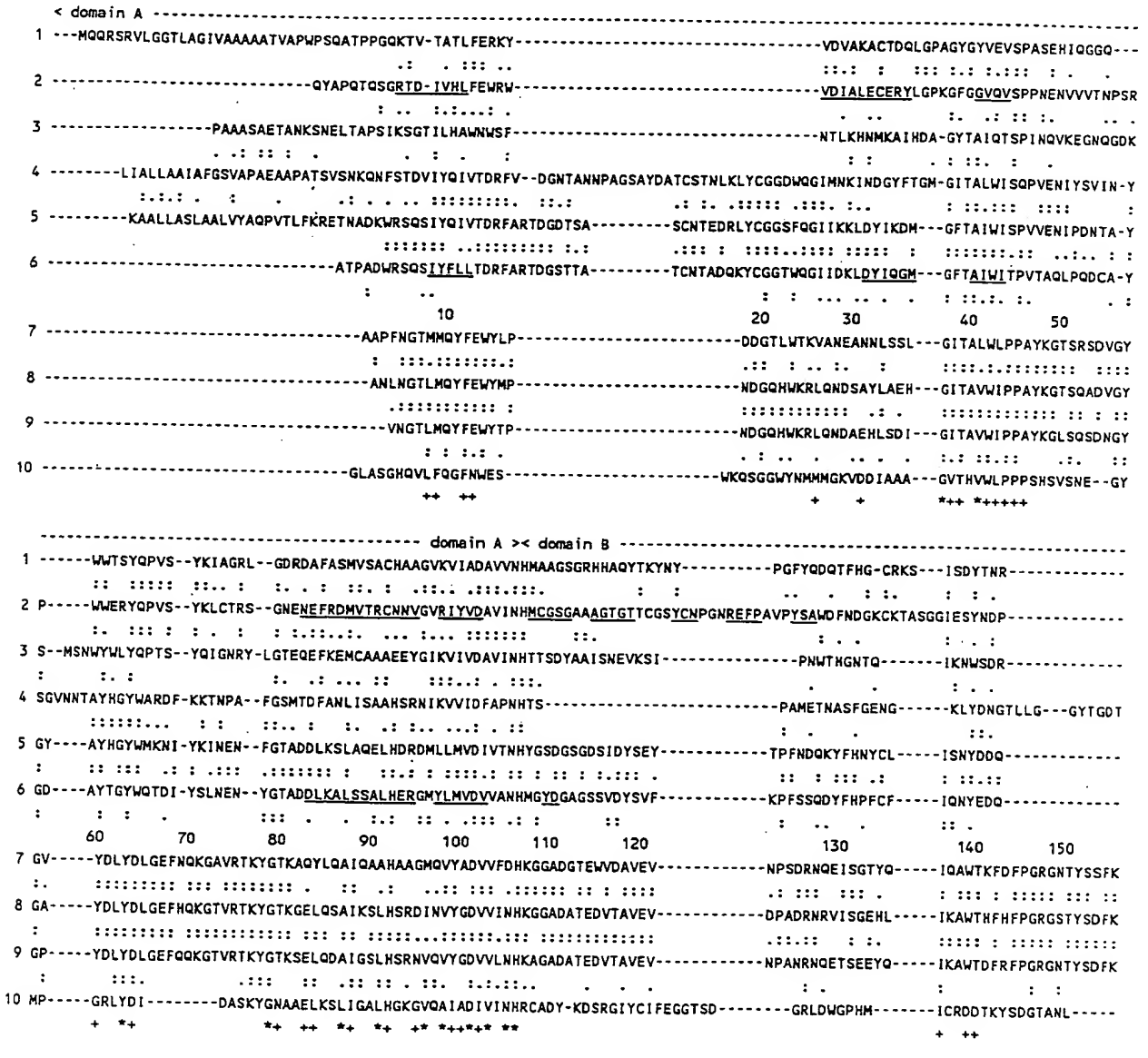

Fig. 1.

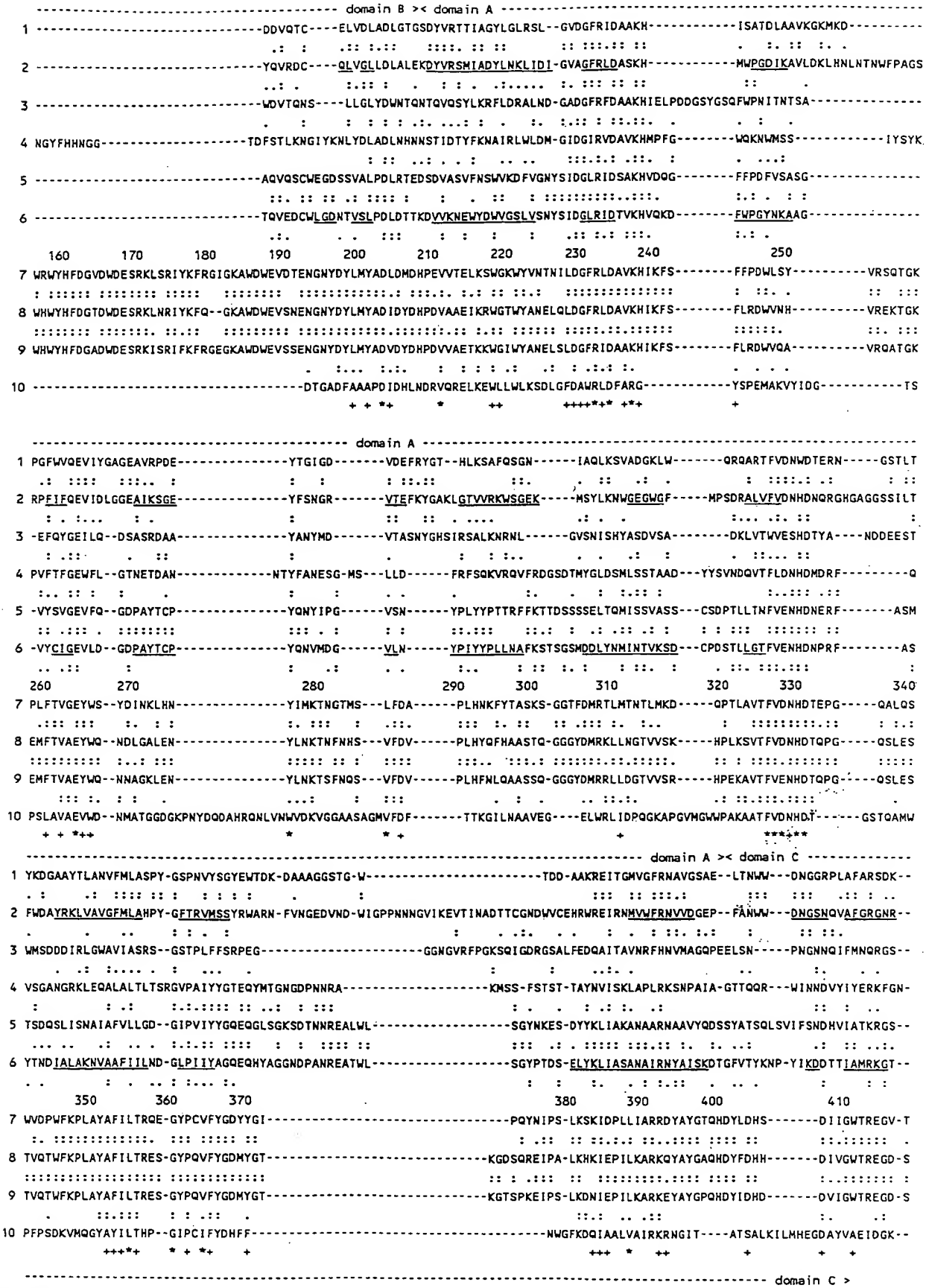

Fig. 1 (continued).

```
1 -----GFVAL---NNGDAALTQTFA--TSLPAGTYCDVV---HAASSCDGD--TVTVG-DTEA-QVDAAKS-VALHVGATGQSACRQAVALHVPGQSAGSPRSSAKRVEQ
2 -----GFIVF---NNDDWQLSSTLQ--TGLPAGTYCDVISGDKVGNSCTGI--KVYVSSDGKA-QFSISNSAEDPFIAIHAE----SKL
3 ----HGVVLA-----NAGSSSVSINTATKLPDGRY-----DNKAGAG------SFQVN-DGKL-TGTINARSVAVLYPDDIAKAPHVFLE
4 -----NVAVVAINKNLTSSYSIAGL-NTSLPSGTYT-----DVLANSLSGN--SITVGSSGAVNTFTLQAGGEASGLTRRRQRLR
5 ------VVSVF----NNLGSSGSSD--VTISNTGYS---SGEDLVEVLTCS--TVSGSSDLQVS---IQGGQ--PQIFVPA--KYASDICS
6 --DGSQIVTILS-NKGASGDSYTLS----LSGASYT---AGQQLTEVIGCT--TVTVGSDGNV-PVPMAGGL--PRVLYPTEKLAGSKICSDSS
  420       430       440                 450       460       470       480       490       500       510
7 EKPGSGLAALITDGPGGSKWMYVGK------------QHAGKVFYDLTGNRSDTVTINSDGWG-EFKVNGGS--VSVWVPRKTTVSTIAWSITTRPWTDEFVRWTEPRLVAWP
8 SVANSGLAALITDGPGGAKRMYVGR------------QNAGETWHDITGNRSEPVVINSEGWG-EFHVNGGS--VSIYVQR
9 SAAKSGLAALITDGPGGSKRMYAGL------------KNAGETWYDITGNRSDTVKIGSDGWG-EFHVNDGS--VSIYVQK
10 ------VVVK---IGSRYDVGAVI------PAG-FVTSAHGNDYAVW-----------------EKNG--AAATLQRS
```

Fig. 1. Sequence alignment of α-amylases from different sources. Aligned α-amylase amino acid sequences from: **1**, *Streptomyces hygroscopicus* SF-1084 strain AA69-4 (Hoshiko *et al.*, 1987); **2**, pig pancreas (Pasero *et al.*, 1986); **3**, *Bacillus subtilis* strain N7 (Yamane *et al.*, 1984); **4**, *B.circulans* (Nishizawa *et al.*, 1987); **5**, *Saccharomyces fibuligera* (Yamashita *et al.*, 1985); **6**, *Aspergillus oryzae* (Matsuura *et al.*, 1984); **7**, *B.stearothermophilus*; **8**, *B.licheniformis* (Yuuki *et al.*, 1985); **9**, *B.amyloliquefaciens* (Takkinen *et al.*, 1983); **10**, barley (Rogers, 1985). Five more mammalian α-amylase sequences are included in NBRF protein database (Barker *et al.*, 1988). They have >80% identical residues with the pig pancreatic α-amylase and were therefore excluded from the alignment. Identical amino acids in adjacent sequences are marked by a colon. A dot marks changes within the following groups: {P, G}, {S, T}, {E, D, N, Q}, {H, Y, F, W}, {A, I, L, V, M, C}, {K, R}. Gaps are denoted by −. An asterisk below the alignment marks positions conserved in all sequences, and a plus denotes positions where variability is limited to two different groups. α Helices and β strands are underlined in the known structures. Gaps correspond to surface loops in the structure of Taka-amylase A. Residue numbers for *B.stearothermophilus* are shown above its sequence.

Bacillus α-amylases, domain B contains a 43−45 residues long insert relative to Taka-amylase A. For these reasons the proposed structure for domain B should be treated with considerable caution.

Model building

The alignment shown in Figure 1 was used to construct a three-dimensional model of *B.stearothermophilus* α-amylase using the structure of Taka-amylase A as a template (Figure 2). Due to the lack of a suitable structural template, two segments have been omitted from the model: the 45-residue insert in domain B—the model joins residue 142 to residue 189—and the C-terminal extension of 19 residues (residues 497−515). Without these two segments, construction of the structural models was relatively straightforward since the remaining shorter insertions and deletions involve only a few residues. Insertions or deletions also occur in places where the polypeptide chain folds back on itself so that loops of different lengths are easy to accommodate. Thus the structural changes made during modelling were minimized. One difficult region was between amino acids 444 and 456, in which a loop between two β strands in the Taka-amylase A structure is shortened by five residues. This deletion could only be incorporated by using a long extended loop which caused concomitant restructuring of this region by displacing one β strand. As a result, domain C appears more flat and open in our model than in Taka-amylase A.

Mutant library

Point mutations distributed throughout the α-amylase gene were generated by an enzymatic misincorporation method (Lehtovaara *et al.*, 1988). In the work presented here, mutants were identified by sequencing and subsequently assayed on starch plates for enzyme activity. In this sample of 98 phenotypic mutants, there are 31 which show no activity, 28 with decreased activity and 39 with wild-type enzymatic activity (Table II). Seventy-five different residues are hit by at least one phenotypic mutation,

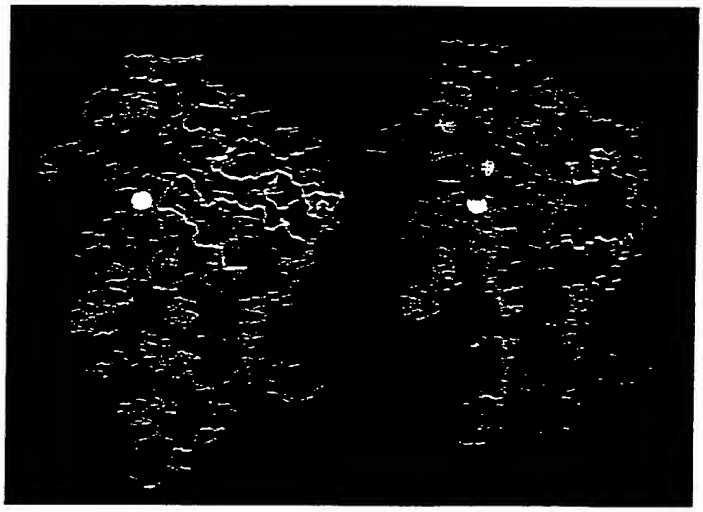

Fig. 2. Structural model of *B.stearothermophilus* α-amylase based on the known structure of Taka-amylase A. Left: Taka-amylase A, right: model of *B.stearothermophilus* α-amylase. Only the N, C(α) and C atoms of the main chain are shown. The model was constructed based on sequence homology to Taka-amylase A, the structure of which has been determined by X-ray crystallography. A 45-residue insert from the back of domain B and a 17-residue extension to the C terminus have been omitted. The light blue sphere shows the location of the essential calcium ion as found in Taka-amylase A. Red, green and yellow spheres mark the probable catalytic residues. Loops 17−38 and 333−350 of Taka-amylase A and the corresponding loops 15−19 and 370−374 of *B.stearothermophilus* α-amylase are coloured yellow. Domain B is coloured orange and domain C green in *B.stearothermophilus* α-amylase. Domain A (blue) is an $(\alpha/\beta)_8$ barrel.

and Figure 3 shows that they are distributed over most of the protein. Clusters of residues at which mutations affect activity are seen around the active site, and at the interface between

Table II. Mutants generated by the random mutagenesis method

					Activity	Observed amino acids
(i) Mutations to stop codons						
K48stop					−	
R126H	**E129stop**(S131S)				−	
E129stop					−	
K141stop					−	
K155stop					−	
K257stop					−	
Y439F	**K442stop**				−	
(ii) Mutations in the active site						
R232W					−	232: R
D234G					−	234: D
H238N					+	238: HG
(V236V)	**H238Y**				−	
H238Y					−	
(L259L)	T261A				++	
T261S	**E264V**				−	261: WIQTCSA
E264D	Y265S				−	264: E
Y265S					++	265: YVWI
Y265F					+	
D331A					−	331: D
T332P	E333D				−	332: TNM
E333A					++	333: EQYP gap
(iii) Mutations in the substrate-binding groove						
Y15F	*D18V*				++	15: YERSK
Y15F	*D18A*				−	18: YDNQ gap
Y15F	*D18V*	D19V	T21S	L22F	+	19: DSC gap
Y15S	*D18V*	D19V	T21P	L22F	+	21: TQG gap
Y15S	*D18A*	D19V	T21P	L22F	+	22: LHWSTD gap
D18V	D19V				+	
D18V	D19A				+	
D18A	D19V	T21S			++	
D18A	D19V	T21P	L22F		+	
S51R	*S53R*				−	51:SPEYVQIA
D54V	(V55V)	*Y57C*			−	53: SAVDNTGH
D54A	(V55V)	*Y57S*	(G58G)		−	54: DCTIQNS gap
D54V	(V55V)	*Y57S*	(G58G)		−	57: RKY gap
Y60F	*D61V*	*Y63F*			−	60: YGW
Y63C					+	61: DRTHWE
K237R					++	63: Y
P334S					++	
Q336L					+	336: QTDSGA gap
Q336H					++	
Q336P					+	
Y369S					+	369: YMFEPR
Y370S					+	370: YFQEWR
(iv) Mutations in or around the interface between domains A and C						
D343A					+	343: DQPNGS
K347N					+	347: KVLIRY
Y351S					+	351: YIVQWN
I354L					+	354: IFVAM
R413G					−	413: RIKQ
R413W	E414A				−	414: EDKRGS
R413W	E414V				−	
E414V					++	
E414D					++	417: TS gap
T417P	*E418A*				+	418: ES gap
(A426A)	**I428F**				−	428: ILFA gap
S435F					−	435: SAYLGW
S435C	*K436E*				+	436: KDSQAFT
(v) Mutations in other buried sites (solvent-accessible surface of wild-type residue <50 Å^2)						
E13V					++	
D19V	T21S				++	

Table II (continued).

				Activity	Observed amino acids
D19V	T21P	(L22L)	T24P	++	
E29D				+	29: EDKA
N32I	**L33F**	S35C		+	32: NYHDMC
L33F				−	33: LIGKTE
L33F	S35C			+	35: SEDAGFTYQ
K48T				++	
E129G				++	
Q254P				++	
I277F				++	
L291F				+	291: LTIFS
G335C				−	335: GFN gap
G335C	A337T			++	337: ASQE
A337T				++	
(vi) Mutations in other exposed sites (solvent-accessible surface of wild-type residue >50 Å^2					
L22F				++	
K25N				−	25: KRMG gap
N31Y				+	31: NAED
N127T				+	127: NLQTWDF
T133S	Y134F			++	
Y134S				++	
Y134F				++	
Q135L				++	
Q137P				++	
A138T				++	
N149I				++	
H208L				++	
T213A				++	
K216E				++	
S217G				++	
S217C				++	
K220R				++	
T225P				+	225: TEDNMI gap
Y250F				+	250: YHAVST
T255P				−	255: TSA gap
K257T				++	
K257Q				++	
K257M				++	
K419N				++	
K419T				−	419: KVA gap
Y439F	K442Q			+	439: YAGIVSQLE
K442N				++	442: KRLSDNQA gap
Q443H				++	

The wild-type amino acid is given on the left and the amino acid it is mutated to is given on the right of the residue number. Deletorious substitutions are printed in **bold**. Note that some of the mutants have multiple substitutions; comparisons with corresponding single mutants sometimes suggest the causative change. In clusters (like the one around residue 20), where the casuative mutations are less clear, the primary suspects are printed in *italics*. Consequently, all other mutations have been interpreted as being neutral. Silent mutations seen only at the DNA level are in parentheses. Enzyme activities of the mutant clones were estimated on starch plates as compared to the wild-type clone: ++ wild-type activity, + less than wild-type but greater than zero, − zero activity. No clone showed drastically elevated activity. Mutants are listed grouped according to their location in the structural model. The solvent-accessible surface area was calculated using the program DSSP (Kabsch and Sander, 1983) from the model of wild-type *B.stearothermophilus* α-amylase. Residues 1−104 and 208−400 belong to domain A, residues 105−207 to domain B and residues 401−515 to domain C. For convenience of comparison to other α-amylase sequences, the right-most column lists for + and − mutants which amino acids are observed in natural α-amylases at given positions. The data for this listing also includes five mammalian sequences (PIR codes ALHUS, ALHUP, ALMSS, ALMSP, ALRTP) which are highly homologous to pig pancreatic α-amylase and which were excluded from Figure 1.

domains A and C. The relationship between the effects of different mutations and their location on the model is discussed in more detail below.

Active site

There are many conserved and buried charged residues around the active site. D331 is relatively accessible to solvent but the other two proposed catalytic residues, D234 and E264, are close neighbours and appear almost buried at the bottom of a pocket. The most striking mutation in this region is the chemically homologous substitution E264D, which is inactive (in a double mutant with Y265S, which as a single mutation has wild-type activity). This strongly suggests that E264 is essential for catalysis. In addition, the mutation D331A on one hand and the mutant D234G on the other are inactive, suggesting that the carboxyl groups of all three, D234, E264 and D331, play a role in the enzymatic mechanism. In the energy minimized model, R232 is in the centre of a network of hydrogen bonds and salt bridges, which may fix the catalytic residues in the correct orientation for catalysis. R232 is connected to D234, D101, H330

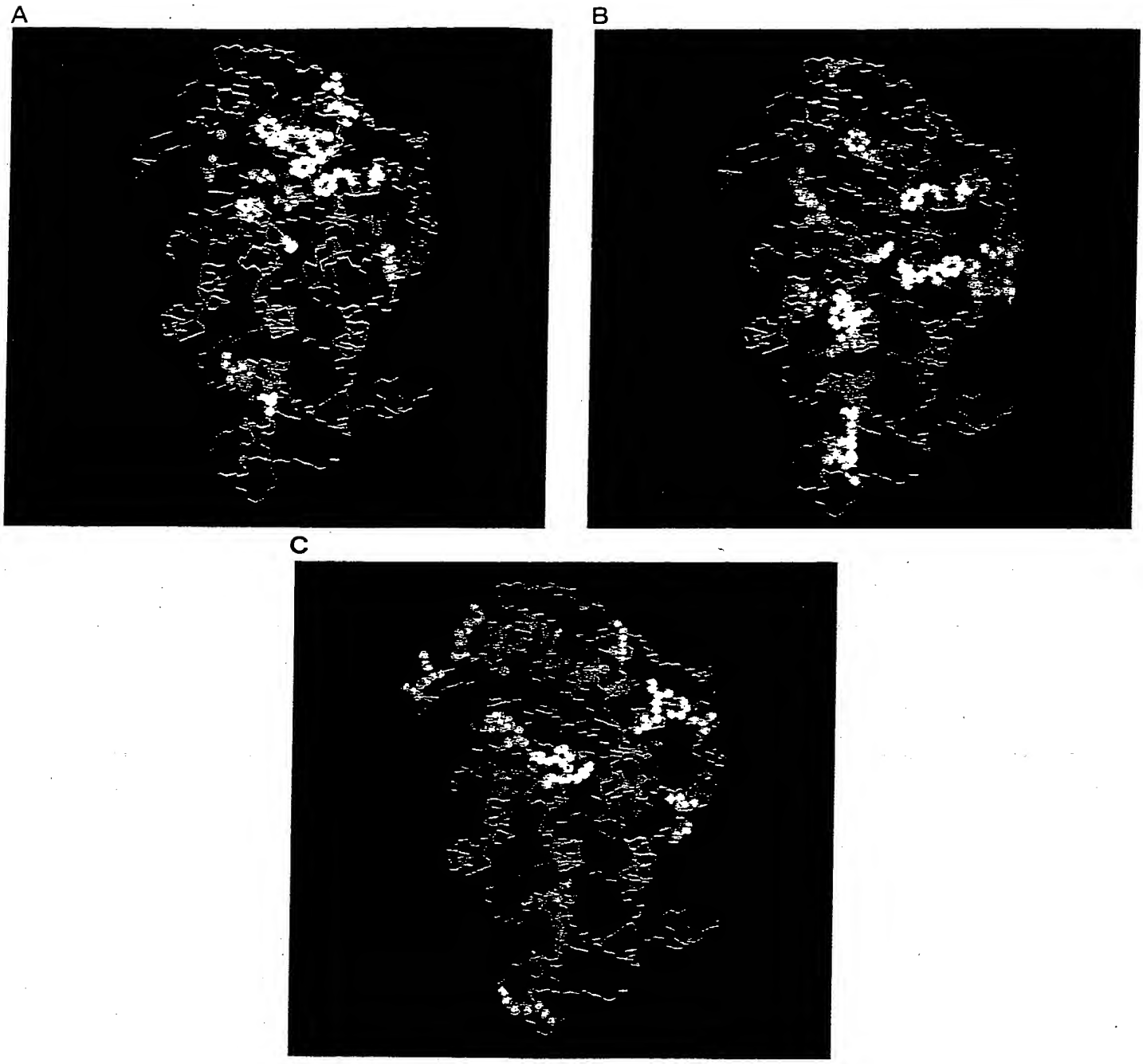

Fig. 3. Location of mutations in the structural model of *B.stearothermophilus* α-amylase. Residues which upon mutation produce (A) loss of activity (− mutants), (B) reduced activity (+ mutants) or (C) wild-type activity (++ mutants). The side chain atoms are drawn as spheres with 0.5 times their van der Waals radius (atom colouring: C green, N blue, O red). Most − and + mutants are clustered around the active site and the interface between domains A and C.

and E264, and further to W42, D331, N329 and D288. All of these residues are highly conserved. In other α-amylases, W42 can be replaced by Q, N329 by S and D288 by N or E or A but all the other mentioned residues are invariant (Figure 1). The mutants R232G and R232W are inactive, thus demonstrating the functional importance of this conserved arginine residue.

Substrate-binding groove

In the model there is a clearly visible groove winding across the top of the α/β barrel in domain A, into which a six-unit amylose chain was docked. Because D234 and E264 are at the bottom of a deep pocket, the long substrate chain could only be brought into contact with them by introducing a sharp bend in the chain. The glucose units make extensive contacts with a number of residues equivalent to those proposed for Taka-amylase A (Matsuura *et al.*, 1984) and porcine pancreatic α-amylase (Buisson *et al.*, 1988). Those residues that are in contact with the substrate at the subsites adjacent to the hydrolyzed bond are highly conserved in all α-amylases, but more variation is allowed at the other subsites.

In Taka-amylase A the region of the substrate-binding groove furthest from the catalytic site is blocked by the adjacent protruding loops 17−38 and 333−350 (residue numbers according to the Taka-amylase A sequence). They correspond to the much shorter loops 15−19 and 370−374 in *B.stearothermophilus* α-amylase (Figure 2). We speculate that the 'missing' subsites might explain the fact that Taka-amylase A is more saccharifying than *Bacillus* α-amylases (Priest, 1984) by biasing the substrate specificity of Taka-amylase A towards residues at the end of the amylose chain.

In the model of *B.stearothermophilus* α-amylase there are notably many tyrosine side chains lining the substrate-binding groove. Several mutations, single or multiple, in these residues (Y15, Y57, Y60, Y63, Y369, Y370) lead to decreased activity (Table II). From its location and mutant data (Table II), Q336 is also implied in substrate binding in *B.stearothermophilus* α-amylase, though it has no counterpart in Taka-amylase A, for example (Figure 1). W266 is in contact with the docked substrate. It corresponds to tryptophan 206 in the barley α-amylase, which has been pinpointed as essential for enzymic function in biochemical mapping (Gibson and Svensson, 1987). Interestingly, in some α-amylases this position is occupied by hydrophobic residues other than tryptophan.

Interface between domains A and C

A second major cluster of inactivating mutations is found in and around the interface between domains A and C, ~30 Å away from the catalytic site (Table II and Figure 3). In low solution X-ray studies of porcine pancreatic α-amylase a second binding site for substrate analogues has been identified near the N-terminus of the eighth helix in domain A that is facing domain C (Payan *et al.*, 1980). The function of this site is unknown, but it has been implied that it may play a role either in anchoring the enzyme to starch or in the regulation of activity (Buisson *et al.*, 1981). In our model both salt bridges and hydrophobic contacts are found across the interface between domains A and C. The latter interactions are affected in the inactive mutant I428F, which is located on a β strand in domain C, and in the mutant I354L, which is on the seventh α helix in domain A and has much decreased activity although this amino acid substitution is conservative. The dramatic effects of such 'innocent-looking' mutations seem to imply that the exact orientation of domain C to domain A is crucial for activity. In the model of *B.stearothermophilus* α-amylase these two residues are opposite to each other. Many combinations of hydrophobic residues are found at these positions in the other α-amylases (Table II, Figure 1). Since, particularly in domain C, many gaps were introduced into the alignment (Figure 1), the structural details of domain C and its contacts with domain A are obviously different in the different enzymes. This is clearly illustrated by comparing the known structures of Taka-amylase A and porcine pancreatic α-amylase (Buisson *et al.*, 1988). That domain C is important for activity is also shown by the effect of a mutation that causes a premature stop codon at residue 442. This removes roughly two-thirds of domain C and leads to inactivation of the enzyme.

Calcium-binding site

The structure of the calcium binding site between domains A and B is similar in the two known structures, and the invariant D203 is one of the ligands of the calcium ion (Buisson *et al.*, 1988). D105 and the main-chain carbonyl oxygen from the invariant H238 are other ligands. D105 is replaced by an asparagine in the other α-amylases (Figure 1). The importance of H-bonds involving the sidechain of H238 is suggested by the mutations H238N, which is still functional, and H238Y, which is inactive (Table II).

Buried sites are typically more sensitive to mutation than exposed sites

The great majority of amino acid replacements which do not much affect the activity of the enzyme are located at exposed sites in the 3-D model. Mutations affecting the active site region and the A−C interdomain region which inactivate α-amylase were described above. Several other examples, where even conservative mutations are deleterious, were found at buried sites. For example, the inactive mutant L33F and the mutant L291F, which has reduced activity, are located in the interior of domain A. Interestingly, the substitution corresponding to L291F is found in *B.circulans* α-amylase. An examination of the structure of Taka-amylase A and models of *B.circulans* and *B.stearothermophilus* α-amylase indicated that Phe can be accommodated at this position in the core of *B.circulans* α-amylase due to a number of compensating substitutions to smaller residues in its neighbourhood, but it could not be fitted to the other two structures without steric clashes.

E29D, located at a partially buried site, has reduced activity. In the three-dimensional model E29 forms a salt bridge to K25. This seems to be an important interaction in the *B.stearothermophilus* enzyme, since the mutant K25N is inactive. The other liquefying *Bacillus* enzymes have D at position 29, but the reduction in size of this residue is compensated by R at position 25. No corresponding salt bridge is present in the structure of Taka-amylase A.

At certain sites proline may be accommodated without effects on the activity, and at other sites it cannot be accommodated (Table II). For example, the mutations Q254P, leading to fully active enzyme, and T255P, leading to inactive enzyme, are in neighbouring residues in a surface loop.

Interactive effects in multiple mutations

Table II shows that generally the effects of amino acid substitutions are additive. There are, however, a number of exceptions. For example, G335C alone gives an inactive enzyme, but activity is recovered if A337T is also present. A337T alone has no effect on activity. Other interesting examples include the clusters of mutations around residues D18 and L33.

Discussion

We describe here the construction of a molecular model of *B.stearothermophilus* α-amylase using a novel alignment of amino acid sequences from many distantly related α-amylases and the available information on secondary/tertiary structures. We also present the phenotype and spatial location of 98 separate randomly generated amino acid mutants and note that inactivating mutations are clustered in the model, revealing functionally important regions. Thus, the model provides a rationale for interpretation of the mutations, and *vice versa*, data on the mutants provides evidence supporting the model.

Porcine pancreatic α-amylase and Taka-amylase A have only 23% identical amino acids (Table I) but very similar three-dimensional structures. We have assumed that all α-amylases, which have typically ~20% identical residues, share this common structural design. There are several well known examples of protein families, such as the globins and the trypsin-like serine proteases, where the sequences can diverge even more but the three-dimensional structures are well conserved. The correct alignment of the amino acid sequences to identify residues

occupying topologically equivalent positions is critical to the success of homology modelling (Read *et al.*, 1984). Argos (1987) has described a cautionary example concerning the pairwise alignment of lactate and malate dehydrogenase: the sequence alignment optimal in terms of amino acid identity does not coincide with the alignment derived from the optimal superposition of the tertiary structures. However, the alignment of α-amylases described here is supported by structural information concerning calcium ligands and secondary structure. In addition, amino acids of similar physical properties are well aligned. At least 70% of residues in all sequence pairs are aligned in Figure 1 and, statistically, the alignment is significant (Table I). It is generally accepted that if the sequence of a protein with unknown tertiary structure can be related to one with known architecture, then the main-chain folding of the latter provides an adequate base for structural modelling (Blundell *et al.*, 1987). However, in the course of evolution, mutations are accommodated in protein structures through small shifts in the relative position of secondary structure elements (Chothia and Lesk, 1982a). Such effects cannot be accounted for in modelling procedures which use a rigid framework. Refinement by energy minimization results only in small shifts in relative atomic positions, typically in the order of 1 Å at most. Thus, it is clear that this model represents only one possible low-energy conformation of the protein. Despite these reservations, we were pleased to note that the α-amylase model is consistent with biochemical data obtained from random mutagenesis.

For this work, it was more important to gather information on a large number of mutants than to analyze a few in detail. Mutants generated by random mutagenesis were identified by sequencing, since it was important to understand what types of base changes the misincorporation procedure produced (Lehtovaara *et al.*, 1988). Around 40% of phenotypic mutants in the sample described here retained wild-type activity. The remainder had reduced activity or were inactive.

The loss of enzyme activity observed in the plate assay could be assayed by: (i) impaired function of secreted enzyme or (ii) a reduction in the amount of enzyme due to destabilization of the secreted protein or impaired folding and even degradation inside *E.coli*. Mutational analysis of a number of proteins has shown that amino acids which are critical to protein stability are generally relatively rigid and inaccessible to solvent in the folded protein whilst substitutions at mobile and exposed sites usually have little effect on protein stability (reviewed by Alber, 1989). Exceptions are made by the introduction of prolines or replacement of conformationally special glycines, which can propagate shifts in the folded structure even if the mutation is at an exposed site (Alber, 1989). Our interpretations concerning the effects of the mutations are based on the location and interactions of the mutated residues in the structural model. Mutations which apparently impair function are clustered together in three-dimensional space at two surface sites: the active site cleft and the domain A–domain C interface. In the substrate-binding groove, a number of exposed residues, which may form hydrogen bonds to the substrate, were shown to be sensitive to mutation. The catalytic center and calcium-binding site are also identified from the known X-ray structures and by sequence conservation. Seven absolutely conserved charged residues in the family of α-amylases are located here. Four of these residues were mutated in our sample and led to inactive enzyme in each case. Apart from the clusters of deleterious mutations mentioned above, exposed residues were found to be less sensitive to mutations than buried ones. A number of mutations to proline and mutations substituting a larger sidechain at buried sites were apparently deleterious because they disrupt the tertiary structure.

The pH dependence of α-amylase suggests a lysozyme-type mechanism employing two carboxylic acids, one ionized and the other protonated in the active form. The adjacent E264 and D234 are buried, and therefore may have elevated pK_as, whereas D331 is more exposed and is expected to have a normal pK_a. Matsuura *et al.* (1984) have argued that in Taka-amylase A two carboxyl groups in different environments are required for activity and therefore the residue corresponding by homology to D331 in *B.stearothermophilus* α-amylase is a catalytic residue. The proposals by Matsuura *et al.* (1984) and Buisson *et al.* (1988) agree on the role of D331 as a catalytic residue in α-amylases but differ with respect to the roles of the neighbouring residues E264 and D234. Buisson *et al.* (1988) reason that D234, which has more direct interactions with the calcium-binding site, is likely to be the other catalytic residue. However, our mutant data very strongly suggest that E264 is a catalytic residue.

Domain C is shown here to be indispensable for activity since a cluster of mutations far from the active site reduce enzyme activity. Subtle conformational changes at the interface between domains A and C, which we assume accompany the mutations I428F and I354L, seem to have a large effect on activity. It has been proposed for mammalian α-amylases that a site at the domain A–domain C interface might have a regulatory role in that substrate analogues bind to this region and affect activity (Payan *et al.*, 1980; Buisson *et al.*, 1981). It is not clear why a bacterial exoenzyme should require this type of regulation. We therefore propose that domain C plays an important role in starch hydrolysis by orientating the active site cleft of domain A correctly with respect to the amylose chain.

In conclusion, we suggest that this approach of integrating computer-aided molecular modelling to the rationalization of random mutagenesis data can contribute substantially to an understanding of protein structure and function, even in cases like this where the crystal structure of that protein is not available.

Acknowledgements

We thank Chris Sander for a helpful comment and Matti Saraste for critical reading of the manuscript. We are grateful to Tuula Teeri and Leif Laaksonen for their momentous input in providing for excellent molecular modelling facilities at VTT and FSCC. Expert technical assistance in making and analyzing the mutants by Kariitta Berg is gratefully acknowledged. This work was supported by the Academy of Finland (S.R.C.) and the Technology Development Centre (TEKES).

References

Alber,T. (1989) *Annu. Rev. Biochem.*, **58**, 765–798.
Argos,P. (1987) *J. Mol. Biol.*, **193**, 385–396.
Barker,W.C., Hunt,L.T., George,D.G., Yeh,L.S., Chen,H.R., Blomquist,M.C., Seibel-Ross,E.I., Elzanovski,A., Bair,J.K., Lewis,M.T., Davalos,D.P. and Ledley,R.S. (1988) National Biomedical Research Foundation, Georgetown University Medical Center, Washington DC.
Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.
Blundell,T.L., Sibanda,B.L., Sternberg,M.J.E. and Thornton,J.M. (1987) *Nature*, **326**, 347–352.
Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) *J. Comp. Chem.*, **4**, 187–217.
Buisson,G., Duee,E., Haser,R., Payan,F. and Darbon,N. (1981) *Acta Crystallogr.*, **A37**, C-34 (suppl.).
Buisson,G., Duée,E., Haser,R. and Payan,F. (1988) *EMBO J.*, **6**, 3909–3916.
Chothia,C. and Lesk,A.M. (1982a) *J. Mol. Biol.*, **160**, 309–323.
Chothia,C. and Lesk,A.M. (1982b) *J. Mol. Biol.*, **160**, 325–342.
Chothia,C. and Lesk,A.M. (1986) *EMBO J.*, **5**, 823–826.
Gibson,R.M. and Svensson,B. (1987) *Carlsberg Res. Commun.*, **52**, 373–379.

Hoshiko,S., Makabe,O., Nojiri,C., Katsumata,K., Satoh,E. and Nagaoka,K. (1987) *J. Bacteriol.*, **169**, 1029–1036.
Jones,T.A. and Thirup,S. (1986) *EMBO J.*, 5, 819–822.
Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.
Knowles,J.R. (1988) *Science*, **236**, 1252–1258.
Lehtovaara,P.M., Koivula,A.K., Bamford,J. and Knowles,J.K.C. (1988) *Protein Eng.*, **2**, 63–68.
Matsuura,Y., Kusunoki,M., Harada,W. and Kakudo,M. (1984) *J. Biochem.*, **95**, 697–702.
Needleman,S.B. and Wunsch,C.D. (1970) *J. Mol. Biol.*, **48**, 443–453.
Nishizawa,M., Ozawa,F. and Hishinuma,F. (1987) *DNA*, **6**, 255–265.
Ogasahara,K., Imanishi,A. and Isemura,T. (1970) *J. Biochem.*, **67**, 65–74.
Payan,F., Haser,R., Pierrot,M., Frey,M., Astier,J.P., Abadie,B., Duee,E. and Buisson,G. (1980) *Acta Crystallogr.*, **B36**, 416–421.
Pasero,L., Mazzei-Pierron,Y., Abadie,B., Chicheportiche,Y. and Marchis-Mouren,G. (1986) *Biochim. Biophys. Acta*, **869**, 147–157.
Phillips,D.C., Sternberg,M.J.E., Thornton,J.M. and Wilson,I.A. (1978) *J. Mol. Biol.*, **119**, 329–351.
Priest,F.G. (1984) In Cole,J.A., Knowles,C.J. and Schlessinger,D. (eds), *Aspects of Microbiology 9: Extracellular Enzymes*. Van Nostrand Reinhold UK, Wokingham, p. 34.
Read,R.J., Brayer,G.D., Jurasek,L. and James,M.N.G. (1984) *Biochemistry*, **23**, 6570–6575.
Rogers,J.C. (1985) *J. Biol. Chem.*, **260**, 3731–3738.
Steer,M.L. and Levitzky,A. (1973) *FEBS Lett.*, **31**, 89–92.
Svensson,B. (1988) *FEBS Lett.*, **230**, 72–76.
Takkinen,K., Pettersson,R.F., Kalkkinen,N., Palva,I., Söderlund,H. and Kääriäinen,L. (1983) *J. Biol. Chem.*, **258**, 1007–1013.
Yamane,K., Hirata,Y.I., Furusato,T., Yamazaki,H. and Nakayama,A. (1984) *J. Biochem.*, **96**, 1849–1858.
Yamashita,I., Itoh,T. and Fukui,S. (1985) *Agric. Biol. Chem.*, **49**, 3089.
Yuuki,T., Nomura,T., Tezuka,H., Tsuboi,A., Yamagata,H., Tsukagoshi,N. and Udaka,S. (1985) *J. Biochem.*, **98**, 1147–1156.

Received on July 10, 1989; revised on October 18, 1989